

Covariates for GLM

KG Katy AM Gaythorpe NF Neil Ferguson

Updated date: Sep 5, 2022

 An abbreviated version of this protocol was published in eLIFE in Mar 2021

The global burden of yellow fever

DOI: 10.7554/eLife.64670

Detailed protocol

Covariates for GLM detailed protocol

Worldclim, NASA's LPDAAC

Data for maximum and minimum temperature; annual precipitation and altitude or elevation was downloaded from worldclim.org. Data for the enhanced vegetation index and middle infrared reflectance was downloaded from Nasa's LPDAAC. These are gridded data available at a range of resolutions, the 30s resolution (approximately equivalent to 1km²) is the most appropriate here. These gridded data were intersected with shapefiles (available from GADM) in order to calculate the mean values over the subnational unit of interest. For example, an elevation is recorded at each point on the grid of the worldclim data, taking the average or mean of the points that lie within the shapefile for one subnational unit provides the mean elevation. The range in temperature is calculated as the difference between max and min temperature per subnational unit.

UNWPP

The United Nations World Population Prospects provides populations sizes per country by 5 year age group (in the most recent revision in 2022, they now provide annual). We use this to extrapolate the annual age group population size per country by interpolating the 5-year, 5-age data using the beers algorithm. To divide this between subnational units, we use the gridded population estimates from Landsat. Similarly to the worldclim data, we intersect the gridded data with GADM shapefiles to arrive at the relative population in each subnational unit. We then scale this relative proportion so the total matches UNWPP and assume the population age structure is the same in each subnational unit.

MODIS

Data for land cover was downloaded from MODIS. This is gridded data and similarly to before, we intersect this with shapefiles available from GADM. However, at this point we calculate the proportion of the subnational unit covered by each land type rather than the max or min.

IUCN redlist

Data for the occurrence of NHP species was downloaded from the IUCN redlist. This is gridded data and similarly to before, we intersect this with shapefiles from GADM. We calculate the proportion of the subnational unit for which the NHP species is recorded and if it is above 10%, we mark that subnational unit as having that species eg. a presence, 1.

Occurrence of *Ae. aegypti* and *Ae. albopictus*

This is taken from the supplementary material of Kraemer et al. 2015 and aggregated to the same subnational granularity as other data sources.

Temperature suitability index

The function and parameter estimates are provided in the public code with Gaythorpe et al 2020, https://github.com/mrc-ide/YF_climateChange/blob/master/FUNCTIONS/temp_suit.R. This function is then applied to the mean temperature as calculated previously.

How to cite: (Readers should cite both the Bio-protocol preprint and the original research article where this protocol was used)

- Gaythorpe, K. and Ferguson, N. (2022). Covariates for GLM. Bio-protocol Preprint. bio-protocol.org/prep1901.
- Gaythorpe, K. A., Hamlet, A., Jean, K., Garkauskas Ramos, D., Cibrelus, L., Garske, T. and Ferguson, N. (2021). The global burden of yellow fever. eLIFE. DOI: [10.7554/eLife.64670](https://doi.org/10.7554/eLife.64670)

Copyright: Content may be subjected to copyright.